

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Analisis regresi merupakan salah satu analisis dalam statistika yang dipergunakan untuk menaksir pola hubungan antara variabel prediktor (X) dan variabel respon (Y). Selain digunakan untuk mengetahui bentuk hubungan antar variabel, analisis regresi juga dapat digunakan untuk peramalan (Hardle, 1990).

Terdapat dua pendekatan dalam analisis regresi, yaitu pendekatan regresi parametrik dan pendekatan regresi nonparametrik. Pendekatan regresi parametrik digunakan jika bentuk kurva regresi diketahui, diperlukan pengetahuan masa lalu tentang karakteristik data yang akan diselidiki, dan data bersifat normal. Jika pola hubungan data membentuk pola linear, maka digunakan pendekatan regresi parametrik linear. Jika pola hubungan data kuadrat, maka digunakan pendekatan regresi kuadratik. Berbeda dengan pendekatan regresi parametrik, dalam regresi nonparaametrik bentuk kurva regresi tidak diketahui, dan data cenderung tidak normal. Kurva regresi nonparametrik hanya diasumsikan *smooth* (mulus) dalam arti termuat didalam satu ruang fungsi tertentu (Budiantara, 2005).

Penelitian yang menggunakan pendekatan regresi sering digunakan untuk menyelidiki satu variabel prediktor dengan satu variabel respon yang disebut analisis regresi univariabel. Jika dalam analisis regresi terdapat satu variabel respon dengan variabel prediktor lebih dari satu, maka regresi tersebut dinamakan regresi multivariabel. Sedangkan analisis regresi yang terdapat dua variabel respon dinamakan analisis regresi birespon.

Keadaan data yang tidak diketahui fungsi regresinya seringkali membuat para peneliti memutuskan untuk menyelesaikan penelitiannya atau menganalisis datanya menggunakan analisis regresi nonparametrik. Beberapa penulis seperti

Hardle (1990), Wahba (1990), Budiantara dan Subanar (1997) menyarankan menggunakan regresi nonparametrik sebagai pendekatan untuk model data, agar mempunyai fleksibilitas yang baik.

Salah satu analisis regresi nonparametrik yang sering digunakan adalah spline. Spline merupakan fungsi polinomial terpotong dalam orde  $k$ , yang mana didalam fungsi tersebut terdapat titik-titik penghubung yang disebut titik knot. Titik knot adalah titik perpaduan bersama dimana terdapat perubahan pola perilaku fungsi atau kurva. Orde dalam fungsi menunjukkan ketinggian derajat polinom fungsi tersebut. Titik knot dan orde inilah yang kemudian akan dipakai untuk menentukan model regresi spline. Spline dapat menggambarkan perubahan pola perilaku dan fungsi pada sub interval, mengatasi pola data yang menunjukkan naik atau turun yang tajam dengan bantuan titik-titik knot serta kurva yang dihasilkan oleh relatif mulus (Eubank, 1998).

Bentuk umum, fungsi spline berderajat  $p$  adalah sembarang fungsi yang dapat disajikan dalam bentuk :

$$f(x) = \sum_{i=0}^p \alpha_i x^i + \sum_{j=1}^K \beta_j (x - k_j)_+^p$$

Dengan fungsi terpotong sebagai berikut :

$$(x - k_j)_+^p = \begin{cases} (x - k_j)_+^p, & x \geq k_j \\ 0, & x < k_j \end{cases}$$

Dimana  $\alpha_i, i = 0, 1, \dots, p$  dan  $j = 1, 2, \dots, K$  merupakan parameter dengan  $K$  titik knot  $k_1, k_2, \dots, k_K$ .

$\alpha_i$  : Koefisien pengganda nilai  $x^i$  terhadap variabel respon.

$x^i$  : Variabel prediktor ke- $i$ .

$\beta_j$  : Koefisien pada variabel  $x^i$  truncated knot ke- $j$  pada spline berorde  $K$ .

$k_j$  : Knot ke- $j$  pada variabel  $x^i$

K : Banyak knot dalam variabel prediktor ke-i.

Spline mempunyai kelemahan pada saat orde tinggi, knot yang banyak dan knot yang terlalu dekat, yang mana salah satu dari ketiga hal tersebut menyebabkan terbentuknya matriks dalam perhitungan yang hampir singular, sehingga persamaan tidak dapat diselesaikan. Tiga kriteria yang harus diperhatikan dalam membentuk model regresi spline, yaitu menentukan orde, menentukan banyak knot, dan menentukan lokasi penempatan knot. Orde untuk model dapat ditentukan berdasarkan pola umum yang terjadi pada data, sedangkan banyaknya knot dan lokasi knot ditentukan berdasarkan perubahan pola di daerah tertentu pada kurva. Nilai GCV (*Generalized Cross Validation*) merupakan nilai yang digunakan untuk memilih model yang terbaik dengan memilih GCV yang optimum.

Salah satu ciri majunya suatu negara adalah dengan majunya pembangunan manusia. Pembangunan manusia berperan penting dalam suatu negara. Faktor-faktor yang berpengaruh terhadap pembangunan manusia, diantaranya pendidikan dan ekonomi. Apabila tingkat ekonomi dan tingkat pendidikan tinggi bisa dikatakan suatu negara telah maju. Berdasarkan isi Human Development Report (HDR) pertama tahun 1990, pembangunan manusia adalah suatu proses untuk memperbanyak pilihan-pilihan yang dimiliki oleh manusia. Diantara banyak pilihan tersebut, pilihan yang terpenting adalah untuk berumur panjang dan sehat, untuk berilmu pengetahuan, dan untuk mempunyai akses terhadap sumber daya yang dibutuhkan agar dapat hidup secara layak (BPS, 2016).

Salah satu komponen yang berpengaruh terhadap Indeks Pembangunan Manusia (IPM) adalah tingkat pendidikan yang diukur dengan angka melek huruf dan rata-rata lama sekolah. Angka Melek Huruf adalah persentase penduduk usia 15 tahun keatas yang dapat membaca dan menulis huruf latin dan atau huruf lainnya. Rata-rata lama sekolah menggambarkan jumlah tahun yang digunakan oleh penduduk usia 15 tahun keatas dalam menjalani pendidikan formal. Di Jawa Barat salah satu provinsi di Indonesia merupakan provinsi yang masih rendah

dalam pembangunan manusia contohnya pendidikan. Hal ini perlu adanya peningkatan agar warga Jawa Barat mendapat pendidikan yang tinggi.

Berdasarkan paparan di atas dalam tugas akhir ini akan dibahas mengenai Regresi Nonparametrik Birespon Spline yang akan diterapkan terhadap tingkat pendidikan di Jawa Barat, yaitu dengan variabel respon (Y) Angka Melek Huruf (Y1) dan Rata-Rata Lama Sekolah (Y2) di Jawa Barat dengan dua variabel prediktor (X), yaitu Rata-Rata banyaknya Anggota Rumah Tangga (X1) dan Pengeluaran Pemerintah Daerah pada Bidang Pendidikan (X2).

### 1.2. Rumusan Masalah

Masalah yang dirumuskan dalam penelitian ini adalah:

1. Bagaimana teori tentang Regresi Nonparametrik Birespon Spline?
2. Bagaimana penerapan Regresi Nonparametrik Birespon Spline terhadap Angka Melek Huruf dan Rata-Rata Lama Sekolah di Jawa Barat?

### 1.3. Batasan masalah

Penelitian ini berpusat pada Angka Melek Huruf dan Rata-Rata Lama Sekolah di Jawa Barat dengan jumlah kota dan kabupaten sebanyak 26 dan menggunakan metode Regresi Birespon Spline dengan *Generalized Cross Validation* (GCV) dalam pemilihan titik knot optimal pada spline linear 1 knot dan spline kuadratik 1 knot.

### 1.4. Tujuan Penulisan

Berdasarkan rumusan masalah di atas, maka tujuan penulisan skripsi ini adalah sebagai berikut:

1. Mengetahui teori tentang Regresi Nonparametrik Birespon Spline.
2. Mengetahui penerapan Regresi Nonparametrik Birespon Spline terhadap Angka Melek Huruf dan Rata-Rata Lama Sekolah di Jawa Barat.

### **1.5. Manfaat Penulisan**

Adapun manfaat yang diharapkan penulis dari tugas akhir ini adalah:

#### **1. Manfaat teoritis**

Menambah wawasan dalam statistika, khususnya mengenai pemodelan regresi nonparametrik salah satunya dengan birespon spline.

#### **2. Manfaat praktis**

Diharapkan dapat menjadi acuan untuk penelitian selanjutnya tentang indeks pembangunan manusia.